THE JOINT DETERMINATION OF PRODUCTION, EMPLOYMENT, AND INVESTMENT DECISIONS

Ray C. Fair

ECONOMETRIC RESEARCH PROGRAM
Research Memorandum No. 128
September 1971

The research described in this paper was supported by NSF Grant GS-2799.

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

September 1971

THE JOINT DETERMINATION OF PRODUCTION, EMPLOYMENT, AND INVESTMENT DECISIONS

Ray C. Fair
Department of Economics, Princeton University

I. INTRODUCTION

It has generally been the case that production, employment, and investment decisions have been analyzed separately rather than within the context of a complete behavioral model of the firm. A noteable exception to this case is the study of Holt, Modigliani, Muth, and Simon [12], which considered the joint determination of production and employment decisions within the context of a cost-minimization model. Also, Lucas [15] has recently postulated a general stock-adjustment model in which the stock of one input may influence the demand for another input, and Nadiri and Rosen [16] have used essentially this model in an empirical study of employment and investment decisions. Coen and Hickman [5] have also worked with a model that takes into account the interrelationship of employment and investment decisions.

The purpose of this paper is to integrate the models of production, employment, and investment decisions developed in [6], [7], and [8] respectively into a complete behavioral model of the firm and then to use this model as a basis for commenting on a number of recent studies of production, employment, and investment decisions. It is the view of this paper that much of the work being done on production, employment, and investment functions suffers from a failure to consider in an adequate way the fact that much of the time firms appear to be operating below capacity and off their production functions.

In addition, it is felt that the common practice of dealing with quarterly, seasonally adjusted, aggregate data has tended to obscure many of the important determinants of the short-run behavior of firms. The paper also questions the usefulness at the present state of knowledge of trying to derive behavioral relationships of the firm from the specification and minimization of a single underlying cost function. In Section II an integration of the models of the three previous studies is undertaken and the empirical results that have been achieved from estimating these models are discussed. In Section III a critique of recent studies is made in light of the model developed in Section II and suggestions for future research are presented.

II. A MODEL OF PRODUCTION, EMPLOYMENT, AND INVESTMENT DECISIONS

Overview

Sales and factor prices are assumed to be exogenous to the firm, and the firm is assumed to minimize costs given the expected paths of sales and factor prices. The firm is assumed to employ a certain number of workers, to own a certain number of machines (including buildings), to have a certain stock of raw materials on hand, and to have a certain stock of finished-goods inventories on hand. The time interval under consideration is taken to be a month, and the key decision variables of the firm are taken to be the amount of output produced during month t, Y_t, the numbers of workers employed during month t, M_t, the average number of hours each worker is paid for during month t, HP_t, and the number of machines purchased during month t. Other decision variables of the firm, which are of less concern here, include the amount of raw materials purchased during month t and the number of machines scrapped during month t.

Obvious costs to the firm are costs of raw materials, costs of paying workers, costs of holding finished-goods inventories, and costs of owning machines. The costs of owning machines are a function of the original prices of the machines, of the costs of borrowing the funds to purchase the machines, and of tax and depreciation laws. Less obvious costs to the firm may include costs of <u>changing</u> the rate of production, of <u>changing</u> the number of workers employed, and of <u>changing</u> the number of hours paid-for per worker. In particular, a firm may be reluctant to allow large fluctuations in its work force because of the direct costs involved in laying off and rehiring workers (e.g., the costs of having a large personnel department, of severance pay, and of unemployment insurance compensation) and because of the difficulty the firm may have in attracting good workers if it has a reputation of poor job security. In addition, such things as guaranteed annual wages and other long-term commitments to some workers may lessen any wage costs that the firm can save by allowing large fluctuations in its work force.

Since the firm is assumed in this study to minimize costs given expected paths of sales and prices, one approach that could be taken here in analyzing the behavior of the firm would be to specify a cost function, differentiate the cost function with respect to the decision variables, and then solve for the equations explaining the decision variables. This is the approach that Holt et al. [12] took in analyzing production and employment decisions and the approach that Childs [4] and Belsley [1] took in analyzing inventory or production decisions. An alternative approach is to specify the equations explaining the decision variables directly, taking into account cost-minimizing considerations but not writing down the actual form of the cost function.

Which of the above two approaches is preferable depends on how well the cost function can be specified. If the cost function can be specified well, then there is no need to use the second, less elegant approach. The view taken here, however, is that the various costs facing the firm and the interrelationship of these costs are too complicated for there to be much chance at the present state of knowledge that a realistic cost function can be specified. Certainly quadratic cost functions, which have been used by Holt et al., Childs, and Belsley, appear to be too simple,[1] especially considering the likely asymmetry of costs near capacity output, and nonlinear, nonquadratic cost functions are difficult to work with. An attempt to specify nonlinear, nonquadratic cost functions may also be based on as much, if not more, theoretical ad-hockery as is involved in the specification of equations explaining the decision variables directly. There is certainly a tendency among economists, as witnessed, for example, in Zvi Griliches' survey article [11] and Marc Nerlove's Schultz lecture [17], to want to derive behavioral rules from first principles -- in most cases cost or profit functions -- but it is not clear that economists know enough yet about the underlying cost or profit functions of the firm for this to be such an all-encompassing goal. It will be argued in Section III that work of a more pedestrian nature is needed before attempts are made to deduce the behavior of firms from a single underlying cost or profit function.

The following, then, is an outline of a model of the behavior of the firm based on direct specification of the equations explaining the decision variables. A more detailed description of the individual models of production, employment, and investment decisions can be found in [6], [7],

---

[1] Evidence is presented in [6] and [7], for example, that rather strongly indicates that the derived production and employment equations of the Holt et al. model are not realistic descriptions of firm behavior. More will be said about this in Section III.

and [8]. The outline here differs from the discussion of the individual models in that much more attention is paid here to the interrelationships among the three kinds of decisions.

The Underlying Technology

The underlying technology is assumed to be of a "putty-clay" type, where at any one time there are assumed to be k different types of machines that can be purchased. The machines differ in price, in the number of workers that must be used with each machine per unit of time, and in the amount of output that can be produced per machine per unit of time. The worker-machine ratio is assumed to be fixed for each type of machine. Let $\lambda_{\nu i}$ denote the amount of output that can be produced per hour per worker on machines of number i that were purchased in month ν, and let $\mu_{\nu i}$ denote the amount of output that can be produced per hour on one of these machines. Also, let $I_{t\nu i}$ denote the number of machines of number i that were purchased in month ν that are actually operating in month t, let $M_{t\nu i}$ denote the number of workers working in month t on these machines, let $H_{t\nu i}$ denote the number of hours worked per worker and machine in month t on these machines, and let $Y_{t\nu i}$ denote the amount of output produced in month t on these machines. The machines will be assumed not to be subject to physical depreciation, so that $\lambda_{\nu i}$ and $\mu_{\nu i}$ are not a function of t. This assumption and the constancy of the worker-machine ratio imply that

(1) $$Y_{t\nu i} = \lambda_{\nu i} M_{t\nu i} H_{t\nu i} = \mu_{\nu i} I_{t\nu i} H_{t\nu i} \quad .$$

Let m_t denote the age of the oldest machine that is being used in month t. Then Y_t, the total amount of output produced in month t, is

$$(2) \qquad Y_t = \sum_{\nu=t-m_t}^{t} \sum_{i=1}^{k} Y_{t\nu i} \quad .$$

In other words, the total amount of output produced in month t is the sum of the amount of output produced on each of the machines that are operating in month t. The parameter k in equation (2) may not be constant through time since more types of machines may be available for purchase at one time than at another, but without loss of generality k can be assumed to be independent of time. What is implied by this assumption is that as a new type of machine becomes available for purchase, one of the old types of machines is phased out. Technical progress is assumed to be reflected in the new types of machines, i.e., in the λ and μ coefficients of the new types of machines.

Equation (1) refers to the actual amount of output produced on a given type of machine, but it can also be written in terms of capacity output. Let $I_{\nu i}$ denote the total number of machines of number i that were purchased in month ν (regardless of whether they are actually operating in month t or not), let $M_{\nu i}$ denote the number of workers that are required to work on the $I_{\nu i}$ machines, and let $H_{\nu i}$ denote the "capacity" number of hours that can be worked per worker per month on these machines. Then capacity output per month on these machines, $Y_{\nu i}$, is

$$(3) \qquad Y_{\nu i} = \lambda_{\nu i} M_{\nu i} H_{\nu i} = \mu_{\nu i} I_{\nu i} H_{\nu i} \; .$$

If m'_t denotes the age of the oldest machine in existence in month t,[2] then total capacity output in month t, Y_t^c, is

$$(4) \qquad Y_t^c = \sum_{\nu=t-m'_t}^{t} \sum_{i=1}^{k} Y_{\nu i} \quad .$$

Likewise, the capacity number of workers in month t, M_t^c, is

$$(5) \qquad M_t^c = \sum_{\nu=t-m'_t}^{t} \sum_{i=1}^{k} M_{\nu i} \quad .$$

Capital-Labor Substitution

Let W_t denote the average wage rate in month t and let C_t denote the average cost of new capital in month t. The cost of capital will be different for each type of machine since the cost of capital is a function of the price and length of life of the capital good in question in addition to being a function of the cost of borrowing funds and of tax and depreciation laws. For purposes here, however, it will be assumed that the prices of the various types of machines, while differing from each other at any one time, move together over time. If this is true, then an average cost of capital can be

[2] For simplicity and without loss of generality, machines of the same age are assumed to go out of existence all at the same time.

defined based on the average price of the various types of machines and on the average length of life of the various types of machines. Since the main concern in this subsection is with how the cost of capital changes relative to the wage rate, the use of the concept of an average cost of capital should not pose any serious difficulties in the present analysis. The ratio W_t/C_t will be referred to as the wage-rental ratio.

If more than one type of machine can be purchased at any given time, then the wage-rental ratio should have an effect on the types of machines purchased. A high wage-rental ratio should cause firms to purchase high-priced machines with low worker-machine ratios, and a low wage-rental ratio should cause firms to purchase low-priced machines with high worker-machine ratios. This effect of the wage-rental ratio can be considered to be a direct form of capital-labor substitution.

Salter has pointed out two other ways in which capital-labor substitution may take place.[3] First, the direction that technical progress takes may be a product of changing factor prices rather than of new knowledge. To the extent that, say, a rising wage-rental ratio causes technical progress to be biased toward labor-saving techniques (i.e., labor-augmenting technical progress growing faster than capital-augmenting technical progress), this can be considered to be a form of capital-labor substitution. Second, capital-labor substitution may come about by speeding up the rate of replacement investment. If, for example, technical progress is on average biased toward labor-saving techniques, then speeding up the rate of replacement investment will result on average in the purchase of machines with lower worker-machine ratios than the machines being replaced.

---

[3] Salter [18], pp. 24, 71.

There are thus three possible ways in which the wage-rental ratio can affect capital-labor substitution. The ratio can affect the types of machines purchased, the direction of technical progress, and the rate of replacement investment. Data on Y_t^c are available for the Cement and Steel industries in the United States, and data on M_t^c can be constructed (at least approximately) for these two industries. Using these data and data on W_t/C_t, a test of capital-labor substitution can be made based on the above model. This test and the results for the Cement and Steel industries are described in [8]. The results in [8] indicate that capital-labor substitution has definitely taken place in the Cement industry and has probably taken place in the Steel industry.

The Optimal Production-Smoothing Plan and the Demand for Capacity

In the present model the firm is conceived of as making two basic kinds of investment decisions, the first regarding the desired level of capacity and the second regarding how many machines to replace and what types of machines to purchase to meet the desired level of capacity. As seen above, capital-labor substitution relates to the second kind of decision: how many machines to replace and what types of machines to purchase should be a function of the wage-rental ratio. This choice is one aspect of the firm's cost-minimizing behavior. The first kind of decision, the demand for capacity, is assumed to be closely related to production-smoothing decisions, and so attention must now turn to the explanation of production decisions.

The basic idea behind the model of production decisions is that because of costs of changing the rate of production, firms are likely to try to smooth production relative to sales. This idea is common to the studies of Holt et al., Belsley, and many others. The present model differs from previous models in the specification of the way that production is smoothed relative to sales and in the assumption that costs of capital may affect production-smoothing decisions. It will be seen that there are two ways in which the model allows for production to be smoothed relative to sales and that costs of carrying capacity should affect production-smoothing decisions.

If the sales of a firm fluctuate, the firm can smooth production relative to sales by the accumulation and decumulation of inventories.[4] How much production is smoothed relative to sales should depend on such things as the costs of holding inventories relative to the costs of changing the rate of production and the costs of carrying capacity. If, for example, production is not smoothed very much relative to sales, then inventory holding costs are low but a lot of capacity is needed to meet production requirements during the peak sales periods, whereas if production is smoothed very much relative to sales, then inventory holding costs are on average high but less capacity is needed to meet the peak production requirements. The firm is assumed here to weigh the costs of holding inventories, the costs of changing the rate of production, and the costs of carrying capacity and to arrive at an optimal production-smoothing plan. The desired amount of capacity is then assumed to be equal to the amount of capacity necessary to meet peak production requirements under the optimal plan, given the expected path of sales. The decision on the desired amount of capacity is thus assumed to be a consequence of the production-smoothing decision and so to be a function of the same costs that influence the production-smoothing decision.

---

[4] From now on "inventories" will refer only to finished-goods inventories.

Given that firms do smooth production relative to sales, the task remains to specify the equation explaining the rate of production that results from the cost-minimizing behavior of the firm. It will first be convenient to introduce two concepts of the desired stock of inventories, a long-run desired stock of inventories, $\bar{V}$, and a short-run desired stock of inventories that ignores information regarding the previous rate of production, V_t^d. If sales were expected to be constant through time, then inventories would really not be needed at all except for such things as insurance against an unexpected increase in sales or breakdown in production, and the desired stock of inventories could be taken to be constant through time. $\bar{V}$ is used to denote this "long-run" or "average" desired stock of inventories.

Since expected sales do fluctuate in the short run, the short-run desired stock of inventories is likely to fluctuate also. If sales are expected to increase over the next few periods, the short-run desired stock of inventories is likely to be larger than $\bar{V}$ so that part of the increase in sales can come from drawing down inventories rather than by increasing production to the full extent of the increase in sales; and if sales are expected to decrease over the next few periods, the short-run desired stock of inventories is likely to be smaller than $\bar{V}$ so that part of the decrease in sales can come from building up inventories rather than by decreasing production to the full extent of the decrease in sales. The difference between the short-run and long-run desired stock of inventories is thus assumed to be a function of expected future changes in sales:

$$(6) \qquad V_t^d - \bar{V} = \sum_{i=1}^{n} \gamma_i (S_{t+i}^e - S_{t+i-1}^e) \quad ,$$

where S^e_{t+i} is the amount expected (at the beginning of month t) to be sold during month t+i. V^d_t is the short-run desired stock of inventories for the end of month t (desired at the beginning of month t).

The decision variable to be explained is the amount of output produced during month t, Y_t. Let S_t denote the level of sales during month t, and let V_t denote the stock of inventories on hand at the end of month t. By definition, production equals sales plus the change in the stock of inventories, $Y_t \equiv S_t + V_t - V_{t-1}$, and so with S_t exogenous, once Y_t is determined, V_t is automatically determined. The desired amount of output produced, Y^d_t, corresponding to expected sales, S^e_t, and V^d_t is

$$(7) \qquad Y^d_t = S^e_t + V^d_t - V_{t-1} \quad .$$

If sales expectations were always perfect and the firm always produced Y^d_t amount of output, then the actual stock of inventories at the end of month t, V_t, would always be equal to V^d_t. In practice, however, the firm is likely to realize that its sales forecasts are not always perfect, and since there are costs of changing the rate of production, the firm may decide to smooth production even more than is implied by producing Y^d_t each period. Remember that Y^d_t is based on V^d_t, which ignores information regarding the previous rate of production. Consequently, a simple lagged adjustment process for planned production is postulated:

$$(8) \qquad Y^p_t - Y_{t-1} = \lambda(Y^d_t - Y_{t-1}) \quad ,$$

where Y^p_t denotes the amount planned to be produced for month t, the plans being made at the beginning of month t. Equations (6), (7), and (8) then imply that

$$(9) \qquad Y_t^p - Y_{t-1} = \lambda\bar{V} + \lambda S_t^e - \lambda Y_{t-1} - \lambda V_{t-1} + \sum_{i=1}^{n} \lambda\gamma_i (S_{t+i}^e - S_{t+i-1}^e) \quad ,$$

which is the basic equation explaining production decisions.[5]

There are thus two ways in which production is smoothed in the present model. One way is through the response of desired production to expected future sales changes, and the other way is through the lagged adjustment equation for planned production. Equation (9) was estimated using monthly data for four three-digit manufacturing industries in [7] under two basic expectational hypotheses and under the assumption that planned production equals actual production. The results strongly indicate that future sales expectations are important in determining short-run production decisions. This is contrary to the results of Belsley [1], who could find no evidence that future sales expectations are important. Belsley's negative results in this regard were attributed in [7] to the use of questionable data. The specification of equation (9) implies that the coefficient of S_t^e should be equal in absolute value to the coefficient of Y_{t-1},[6] and the hypothesis that this is so was

---

[5]Implicit in the above model is also an equation explaining planned inventory investment:

$$V_t^p - V_{t-1} = \lambda(V_t^d - V_{t-1}) + (1-\lambda)(Y_{t-1} - S_t^e) \ ,$$

where the planned stock of inventories, V_t^p, equals $Y_t^p - S_t^e + V_{t-1}$. This equation is not of direct concern here, however, since inventory investment is assumed to be a consequence of production decisions rather than production being a consequence of inventory decisions.

[6]The specification also implies that the coefficient of V_{t-1} should be the same as the coefficient of Y_{t-1}, but as pointed out in [7], footnote 6, one cannot expect to test this hypothesis from the available data.

tested in [7] and accepted for three of the four industries. This fact and the fact that the inventory variable, V_{t-1}, was more significant when equation (9) was estimated than when Belsley's equation was estimated led to the conclusion in [7] that the present model is an improvement over Belsley's model. The Holt et al. production equation was also estimated in [7] and found not to be realistic.

It has been pointed out that desired capacity should be equal to the amount of capacity necessary to meet peak production requirements under the optimal production-smoothing plan, given the expected path of sales, but so far no equation explaining desired capacity has been specified. One possible approach in this regard would be to estimate equation (9), use the estimated equation to solve for the time path of planned production (given initial conditions and a time path of expected sales), and use the peaks of the planned production series as values of desired capacity. For the work in [8], however, a cruder approach was followed: desired capacity was assumed to be proportional to the level of expected peak sales. This assumption does not appear to be unreasonable since the larger is the level of expected peak sales, the more capacity will be needed to meet peak production requirements under the optimal plan. Assume that the decision period regarding capacity decisions is a year, and let Y_τ^c denote capacity output at the end of year τ, Y_τ^{c*} desired capacity output for the end of year τ, and SP_τ^e the expected level of sales of some peak period of year τ. Then it was assumed in [8] that

$$(10) \qquad Y_\tau^{c*} = \alpha SP_\tau^e \qquad ,$$

and that Y^c_τ was subject to a simple lagged adjustment process:

$$(11) \qquad Y^c_\tau - Y^c_{\tau-1} = \delta(Y^{c*}_\tau - Y^c_{\tau-1})$$
$$= \delta(\alpha SP^e_\tau - Y^c_{\tau-1}) \quad .$$

Equation (11) was estimated in [8] for the Cement and Steel industries under various assumptions about SP^e_τ, and it definitely appeared to be the case that expected peak sales were what determined desired capacity and not, for example, expected sales of the whole year. The results in [8] were thus consistent with the above hypothesis about desired capacity decisions.

It should be noted that if actual capacity is adjusted toward desired capacity with a lag, as in equation (11), then the production-smoothing plan will have to take this into account. It should not be the case that planned peak production for the year is greater than planned capacity. In the above model, desired capacity is meant to refer to that amount of capacity desired assuming no costs of adjustment in changing capacity, and if there are costs of changing capacity, then the firm will have to take this into account when formulating its production and investment plans. Therefore, the firm is likely to weigh both the costs involved in changing the rate of production and the costs involved in changing capacity and jointly decide on the production-smoothing plan for the year and on the investment plan.

Employment Decisions

The capacity number of workers, M^c_t, is a function of capacity output and of the types of machines that are on hand. This can be seen from equations (3), (4), and (5). Since the wage-rental ratio affects the types of machines that are purchased, the wage-rental ratio affects M_t^c. What is of concern in this section, however, is what determines the actual number of workers employed, M_t, and the average number of hours paid for per worker, HP_t. From equations (1) and (2), the number of man hours required to produce output Y_t, say M_tH_t, is

$$(12) \qquad M_tH_t = \sum_{\nu=t-m_t}^{t} \sum_{i=1}^{k} \frac{Y_{t\nu i}}{\lambda_{\nu i}} \quad .$$

If it is assumed that there are no completely idle workers, then the total number of workers employed is equal to the number of workers working on the operating machines:

$$(13) \qquad M_t = \sum_{\nu=t-m_t}^{t} \sum_{i=1}^{k} M_{t\nu i} \quad .$$

Therefore, the average number of hours worked per worker is

$$(14) \qquad H_t = \frac{M_tH_t}{M_t} \quad .$$

The number of hours paid-for per worker can never be less than the number of hours worked per worker, and a key assumption of the present model is that the number of hours paid-for per worker is usually greater than the number of hours actually worked per worker except during peak output periods. If this is true, then it means that in practice the true labor inputs are not observed since the data available on hours worked per worker are actually data on hours paid-for per worker. As discussed in [6], this assumption about hours paid-for per worker provides an explanation of the commonly observed phenomenon of increasing returns to labor services in the short run, and it also implies that the properties of the short-run production function cannot be estimated from the available data.

The present model of employment decisions relies heavily on the idea that there are costs involved in changing the size of the work force in the short run. Some of these costs were mentioned at the beginning of this section and others are discussed in [6], p. 39. It was argued in [6] that there are even likely to be costs involved in changing the number of hours paid-for per worker as well as costs involved in changing the number of workers employed. Again, a firm may have difficulty in attracting good workers if it has a reputation of allowing large fluctuations in the work week.

The standard number of hours of work per worker, denoted as HS_t, is assumed to be the number of hours the firm would like workers to work and be paid for if there were no problems of fluctuating man-hour requirements. HS_t is the dividing line between straight-time hours and more costly overtime hours.[7] Given HS_t, the desired number of workers employed, say M_t^d, corresponding to man-hour requirements M_tH_t is

$$(15) \qquad M_t^d = \frac{M_tH_t}{HS_t} \quad .$$

[7]In some firms a certain amount of overtime work has become standard practice, and for these firms HS_t should be considered to be the standard number of hours of work per worker plus this standard or "accepted" number of overtime hours of work per worker.

M_t^d is the desired number of workers employed in the sense that if man-hour requirements were to remain at the level M_tH_t, M_t^d can be considered to be the number of workers the firm would want to employ in the long run. The concept of excess labor plays an important role in the present model. The amount of (positive or negative) excess labor on hand during month t is defined to be the (logarithmic) difference between the actual number of workers employed and the desired number: $\log M_t - \log M_t^d$. From equations (14) and (15) it can be seen that this measure is also equal to the (logarithmic) difference between the standard number of hours of work per worker and the actual number of hours worked per worker: $\log HS_t - \log H_t$.

The two basic factors that are assumed to affect the firm's decision on how many workers to hire or lay off are the time stream of expected future output and the amount of excess labor on hand. If, for example, output is expected to increase over the next few months, the firm may be reluctant to lay off workers it does not actually need at present, and it may begin to build up its work force in anticipation of higher future man-hour requirements. Conversely if output is expected to decrease over the next few months, the firm may be less reluctant to lay off workers, and it certainly has no need to build up its work force any further. With respect to excess labor on hand, one would expect that the larger the amount of positive excess labor on hand at the beginning of the month, the larger would be the number of workers who would be laid off during the month. Holding positive amounts of excess labor is costly, and the firm can be considered to be continually trying to eliminate this excess labor in the light of adjustment costs and expected future man-hour requirements. Conversely, if there is negative excess labor on hand (too few workers employed), the firm can be considered to be continually trying to add workers to achieve a zero amount of excess labor.

The following was taken in [6] to be the basic equation explaining the short-run demand for workers:

$$(16) \qquad \log M_t - \log M_{t-1} = \alpha_1(\log M_{t-1} - \log M^d_{t-1}) + \sum_{i=1}^{m} \beta_i(\log Y_{t-i} - \log Y_{t-i-1})$$
$$+ \gamma_o(\log Y^e_t - \log Y_{t-1}) + \sum_{i=1}^{n} \gamma_i(\log Y^e_{t+i} - \log Y^e_{t+i-1}) ,$$

where Y^e_{t+i} is expected (or planned) output for month t+i, the expectations being made at the beginning of month t. The past output changes were added to the equation on the argument that they may help depict the reaction of the firm to the amount of excess labor on hand, but the variables were never really very important in explaining $\log M_t - \log M_{t-1}$.

An equation explaining the short-run demand for hours paid-for per worker was also developed in [6]. If there are costs involved in changing the number of hours paid-for per worker, then it seems likely that the same factors that influence the short-run demand for workers will also influence the short-run demand for hours paid-for per worker, and this is the basic premise upon which the equation explaining the short-run demand for hours paid-for per worker in [6] was based. There is, however, one main difference between hours paid-for per worker and workers. Unlike the number of workers employed, which can move steadily upward or downward over time, the number of hours paid-for per worker fluctuate around a relatively constant level of hours. If the number of hours paid-for per worker is not equal to this level, this should, other things being equal, bring forces into play causing it to decline back to this level. Therefore, the (logarithmic) difference between the number of hours paid-for per worker and the standard number of hours of work per worker, $\log HP_{t-1} - \log HS_{t-1}$, was included in the equation explaining the demand for hours paid-for per worker. The basic equation explaining the demand for hours paid-for per worker was thus taken to be

$$(17) \qquad \log HP_t - \log HP_{t-1} = \alpha_1'(\log M_{t-1} - \log M_{t-1}^d) + \alpha_2'(\log HP_{t-1} - \log HS_{t-1}) + \sum_{i=1}^{m} \beta_i'(\log Y_{t-i} - \log Y_{t-i-1}) + \gamma_0'(\log Y_t^e - \log Y_{t-1}) + \sum_{i=1}^{n} \gamma_i'(\log Y_{t+i}^e - \log Y_{t+i-1}^e) \quad .$$

Before equations (16) and (17) could be estimated in [6], a measure of M_t^d had to be found. HS_t was assumed to be a smoothly trending variable, so that from equation (13) measuring M_t^d only required measuring M_tH_t, and this was done as follows. The available data on output per paid-for man hour, Y_t/M_tHP_t, was plotted over time, and at the peaks of the series it was assumed that hours paid-for per worker equaled hours worked per worker, so that one had observations on output per man hour, Y_t/M_tH_t, at the peaks. Now, it can be seen from equations (1), (2), and (5) that the ratio Y_t/M_tH_t is a function of the types of machines in use. If new types of machines with new $\lambda_{\nu i}$ coefficients replace older types of machines, then the ratio will change, and if with an existing stock of machines the percentage of time that each machine is operated changes, then the ratio will also change. The assumption was made in [6] that the ratio changed linearly between peaks, and from this assumption and the available data on Y_t, a series on M_tH_t could be constructed for each period.

This procedure of measuring M_tH_t and thus of measuring the amount of excess labor on hand allows the model to handle the effects of technical progress (as reflected in the new types of machines) on employment demand. If Y_t/M_tH_t is increasing through time, then, other things being equal, M_tH_t and so M_t^d will be falling. The amount of excess labor on hand will thus be increasing. The effects of the growth of technology on employment decisions are thus taken care of by the reaction of firms to the amount of excess labor on hand. If during low output periods firms tend to curtail the use of the less efficient machines more than the use of the more efficient ones, then the assumption that Y_t/M_tH_t moves linearly between peaks will not be realistic, but the possibility that Y_t/M_tH_t moves nonlinearly between peaks was not explored in [6]. With only data on Y_t, M_t, and HP_t available, it is not clear that one could discriminate very easily among different nonlinearity assumptions.

Equations (16) and (17) were estimated in [6] under two basic expectational hypotheses for seventeen three-digit manufacturing industries, and the results strongly indicated that both expected future output changes and the amount of excess labor on hand are significant factors in the explanation of $\log M_t - \log M_{t-1}$ and $\log HP_t - \log HP_{t-1}$. Much of the work in [6] was concerned with testing various hypotheses relating to equations (16) and (17) and of combining the two equations to derive an equation explaining the demand for total man hours paid for, but this aspect of the work in [6] will not be discussed here. The Holt et al. employment equation was tested in [6] and found not to be realistic.

The Relationship Between Production and Employment Decisions

In the above model, decisions regarding the current rate of production are assumed not to be influenced by the number of workers on hand nor by the level of hours paid-for per worker. This does not imply, however, that labor costs have no influence on production decisions. Assume, to take an extreme case, that it were not possible during the course of, say, a year for the firm to vary either the number of workers employed, M, or the number of hours paid-for per worker, HP. Then M·HP would have to be chosen at the beginning of the year to be equal to the peak man-hour requirements of the year. If man-hour requirements fluctuated a lot during the year, then the firm would end up paying for a lot of nonproductive time. Costs of labor in this case are like costs of carrying capacity. Unless inventory costs were very high, the firm would probably decide in this case to smooth production a lot relative to sales. In general, of course, man hours paid for can fluctuate, but at a cost, and this cost should have a significant influence on the production-smoothing plan of the firm. Indeed, the major factor determining the production-smoothing plan may be how costly it is for the firm to vary the number of man hours paid for.

The costs of changing man-hours paid for should thus affect both production and employment decisions. These and other costs cause production to be smoothed relative to sales and employment to be smoothed relative to production. The employment plan is constrained by the restriction that man hours paid for must be at least as great as man hours worked, and what appears to be the case, from the evidence in [6], is that the employment-smoothing policies of firms result in man hours paid for being equal to man hours worked only during peak production periods. Most of the time, then, the firm could produce more output if it wanted to with the same number of man hours paid for, but producing more output would lead to higher inventory costs and cause production to deviate from the optimal plan. Since production is not constrained except during peak periods by man hours paid for, there is thus no reason why the current rate of production should be influenced by the number of workers on hand or by the level of hours paid for per worker.

During peak production periods, the firm is operating on its production function so to speak, and it may be the case that the behavior of $\log M_t - \log M_{t-1}$ and $\log HP_t - \log HP_{t-1}$ is different from that implied by equations (16) and (17) during peak production periods. The hypothesis that the behavior of $\log M_t - \log M_{t-1}$ and $\log HP_t - \log HP_{t-1}$ is different from that implied by equations (16) and (17) during peak production periods was tested in [6], but no evidence that the hypothesis is true could be found. An equation like (16) was also estimated in [6] using expected future sales in place of expected future production. The results using sales data were much poorer than the results using production data. This is, of course, as expected since if the above model is true, employment is smoothed only indirectly relative to sales but directly relative to production.

The Effects of the Wage Rate and the Cost of Capital on the Behavior of the Firm

There are a number of ways in which the wage rate and the cost of capital can affect the various decisions of the firm. First, the wage-rental ratio can affect the types of machines purchased, the direction of technical progress, and the rate of replacement investment. These are the direct effects of capital-labor substitution. Second, the cost of capital can affect the production-smoothing plan of the firm. The higher are the costs of carrying capacity, other things being equal, the more should production be smoothed relative to sales. If the cost of capital affects the production-smoothing plan, then this means that it also affects the desired amount of capacity. Note that this effect of the cost of capital on desired capacity has nothing to do with capital-labor substitution. Third, the adjustment parameter δ in equation (11) determining the actual change in capacity may be a function of the deviation of the cost of capital from some expected long run value. If, for example, the cost of capital is particularly high one year and the firm expects that it will be lower next year, then the firm may postpone some of its planned expenditures until the following year. Similarly, the firm may speed up its planned expenditures if the cost of capital is particularly low one year and is expected to be higher next year. This effect of the cost of capital on capacity investment is purely a matter of timing and also has nothing to do with capital-labor substitution. Fourth, the wage rate may affect the demand for workers and for hours paid for per worker other than through the effect of the wage-rental ratio on man-hour requirements and thus on excess labor. Holding positive amounts of excess labor is costly, and if the wage rate is rising relative to other costs, the firm may be induced to hold on average less excess labor and to allow larger fluctuations in man hours paid for. The possible effect of the wage rate in equations (16) and (17) was not tested in [6] because of lack of adequate data on wage rates. Note that this possible effect of the wage rate on employment demand also has nothing to do with capital-labor substitution.

It should also be noted that the size of the coefficients in the production-smoothing equation (9) depend on the costs of the firm. If capital costs, inventory costs, costs of changing the rate of production, costs of changing capacity, and costs of changing man hours paid for change over time, this will cause the optimal production-smoothing plan to change and thus will cause the coefficients in equation (9) to change. If data on these various costs were available, one could try to incorporate the costs into the specification of equation (9) directly, but as it stands lack of data prevents much from being done along these lines. The specification of equation (9) should be adequate if the various costs are not changing very much relative to each other; otherwise the best that can be hoped for is that the estimated equation picks up the average behavior of the firm over the sample period. Likewise, if the costs of changing man hours paid for change over time, this will cause the coefficients in equations (16) and (17) to change.

III. CRITIQUE OF PREVIOUS STUDIES AND SUGGESTIONS FOR FUTURE RESEARCH

Critique of Previous Studies

The discussion in this section is based on the premise that the above model is a fairly accurate representation of the way firms actually behave with respect to production, employment, and investment decisions. If this premise is true, then the approaches of a number of studies can be criticized for leading to unrealistic conclusions about the behavior of firms.

Consider first the quadratic cost-minimizing approach of Holt et al. [12], Childs [4], and Belsley [1]. Because the firm is subject to various capacity restrictions, it does not seem likely in many cases that quadratic costs will be an adequate approximation to the actual costs of the firm. The firm cannot behave in a symmetric way around capacity output. It is easy to produce below capacity but much more difficult (in fact impossible if capacity output is defined rigorously to mean maximum possible output) to produce above capacity. Since only during peak production periods does it appear that the firm is utilizing all of its capacity and not paying for more hours than are required to produce the output, it does not seem likely that the actual costs that the firm is minimizing are anywhere close to being quadratic. It was argued in [6], p. 115, for example, that the Holt et al. quadratic approximation to overtime costs is likely to be very poor. The approximation is not good if production falls to a low level relative to the work force, and it definitely appears to be the case that production does fall to a low level relative to the work force during much of the year. This poor approximation may be one of the main reasons why the Holt et al. employment and production equations gave such poor results when estimated in [6] and [7]. In some situations, of course, quadratic costs may be a good approximation to actual costs, but in general the above model and results indicate that the quadratic cost-minimizing approach is not likely to be a useful one in analyzing the production, employment, and investment behavior of the firm.

Consider next the approach of Nadiri and Rosen [16]. Nadiri and Rosen assume that in the long run the firm minimizes costs subject to a Cobb-Douglas production-function constraint. The cost function is the sum of wage costs (the wage rate times total man hours paid for), other labor costs (the "user cost of labor" times the number of workers employed), and capital costs (the user cost of capital times the stock of capital). The Cobb-Douglas production function relates output to four inputs: the number of workers employed, the number of hours paid-for per worker,[8] the stock of capital, and the rate of utilization of capital. Nadiri and Rosen differentiate the cost function

---

[8]Nadiri and Rosen do not distinguish between hours paid-for and hours worked and implicitly assume that the two are equal. The hours variable that they use in their empirical work is an hours paid for variable.

with respect to the four input variables subject to the production-function constraint and obtain "optimum long run demand functions".[9] The optimum number of workers employed and the optimum stock of capital are functions of the level of output and of relative factor prices. Nadiri and Rosen then postulate that the firm adjusts the actual input levels toward the optimum levels by means of the Lucas stock-adjustment model. In the Lucas model [15] the rate of adjustment of one factor is allowed to be a function of the level of another factor. In the final equations estimated each input is a function of the current level of output, of the wage-rental ratio, of a time trend, and of the four lagged inputs.

If the model in Section II is a fairly accurate representation of the way firms behave, then the Nadiri and Rosen model does not appear to be realistic. First, sales are exogenous to the firm, not production, and the firm is likely to try to smooth production relative to sales. Second, there appears to be no reason why the demand for capital should be a function of the number of workers on hand and the number of hours paid for per worker, as Nadiri and Rosen's model implies. The demand for capacity is a function of the same factors that influence the production-smoothing plan, and the rate of replacement investment and the types of machines purchased to meet capacity requirements are functions of the wage-rental ratio. It seems very unlikely that either the number of workers employed or the number of hours paid for per worker would influence investment decisions. Indeed, it was argued in Section II that the number of workers employed and the number of hours paid for per worker are not even likely to influence production decisions. Third, the Nadiri and

---

[9] Nadiri and Rosen [16], p. 460.

Rosen model ignores most of the dynamics of employment decisions. The model does not capture in an adequate way the fact that employment is smoothed relative to production in the short run. Finally, as mentioned above, the Nadiri and Rosen model implies that firms never pay for more hours than are actually worked (or in the terminology of Nadiri and Rosen that "firms are on their production functions at every moment of time" with respect to the observed inputs[10]). The evidence in [6] strongly indicates that this assumption is just wrong. Firms appear to spend much of the time off of their production functions with respect to the observed inputs.

It is also of interest to note the difference in the way the present model handles the effects of the wage-rental ratio and the changing composition of the capital stock on the demand for workers from the way that Nadiri and Rosen's model handles the effects. In the Nadiri and Rosen model the wage-rental ratio and the stock of capital enter directly in the equation explaining the number of workers employed. In the present model the desired number of workers employed, M_t^d, is a function of man-hour requirements, M_tH_t, which from equation (12) is a function of the amount of output produced and of the types of machines on hand. If new types of machines are added or if new types of machines replace old types of machines -- the choice of the types of machines being a function of the wage-rental ratio -- then man-hour requirements per unit of output, Y_t/M_tH_t, will change, which will in turn change the amount of excess labor on hand. The changing composition of the capital stock thus affects the demand for workers through its effect on excess labor.

Nadiri and Rosen point out that most existing models of employment and investment demand are special cases of their model. This is true, for example, of the neoclassical investment models of Jorgenson [13], Bischoff [2], and

[10] Nadiri and Rosen [16], p. 461.

The common practice of using quarterly, seasonally adjusted, aggregate data may be one of the reasons why most models do not appear to be realistic descriptions of the behavior of the firm. Most of the conclusions reached in this study about the behavior of the firm were obtained by examining monthly, seasonally unadjusted data at the three-digit industry level. Much of the dynamics of the behavior of the firm is not likely to show up in quarterly, seasonally adjusted data; and aggregating across a large number of industries may also tend to obscure the behavior of the firm. The present model, for example, relies heavily on the distinction between peak and nonpeak periods, and in practice it is much more difficult to distinguish between these two periods using seasonally adjusted quarterly data than it is using seasonally unadjusted monthly data.[13]

Suggestions for Future Research

There has recently been a growing interest in trying to deduce the lag behavior of the firm from first principles, i.e., from the minimization or maximization of a cost or profit function.[14] Deducing the behavior of the firm in this way Nerlove argues avoids the ad hoc nature that characterizes the specification of most distributed lags.[15] As mentioned above, it is the view of this paper that the approach of trying to deduce the behavior of the firm from a single underlying cost or profit function is at present not likely to be a very fruitful one. Much more information about the various cost parameters

---

[13]See [6] and [7] for further discussion of the advantages of using seasonally unadjusted data.

[14]See Nerlove [17] for a discussion of recent work in this area.

[15]Nerlove [17], p. 46.

of the firm is needed before this approach is likely to result in realistic implications about the behavior of the firm. The following are a few suggestions for future research where it is felt that progress can be made in understanding the behavior of the firm.

First, more work is needed in trying to determine what the underlying technology facing the firm is like. How much choice, for example, does a firm actually have when purchasing new machines (in other words, how large is k in the above model), and how rapidly does technical progress render old machines completely obsolete? Work is also needed in trying to determine the relative importance of the three ways in which capital-labor substitution can take place. Salter's study [18] is an important contribution in this area, and more work should proceed along these lines.

Second, work is needed in trying to determine what the primary costs of the firm are. It was argued in Section II, for example, that costs of changing man hours paid for may be the most important costs influencing production-smoothing and employment-smoothing decisions, and attempts should be made to see if this is in fact the case. By examining the importance of the various costs facing the firm, one should be able to improve upon the specification of the equations explaining the decision variables of the firm. The coefficients of the production-smoothing equation (9), for example, are a function of inventory costs, capacity costs, costs of changing the rate of production, and costs of changing capacity, and a better understanding of these costs should lead to an improved specification both of equation (9) and of equation (11) explaining the change in capacity. The specification of the employment equations (16) and (17) should also be able to be improved by a more detailed knowledge of the cost structure of the firm.

Third, the present study has ignored the distinction between production-to-stock decisions and production-to-order decisions and has essentially concentrated only on production-to-stock decisions. These two kinds of decisions are not the same, as Childs [4] and Belsley [1] have emphasized, and it would be useful to add production-to-order decisions to the above theoretical framework. It might also be useful to consider the case in which more than one type of input is produced by the firm and the case in which there is more than one kind of labor.[16] This study has also assumed that output prices are exogenous, and it might be useful to drop this assumption and concentrate on profit-maximizing behavior rather than on cost-minimizing behavior. Questions of uncertainty have only been implicitly considered in this study -- uncertainty enters essentially through the expectation variables -- and perhaps a more explicit consideration of uncertainty in the above framework should be undertaken. Certainly more work needs to be done on trying to determine how expectations are formed.

Fourth, attempts should be made to see if the wage rate is a significant explanatory variable in an equation like (16). It was argued above that the wage rate may have a significant effect in equation (16) even though this has nothing to do with capital-labor substitution. Attempts should also possibly be made to see if better measures of excess labor can be found. It may be, for example, that the use of linear peak to peak interpolations in the measurement of excess labor can be improved upon by the use of nonlinear interpolations.

---

[16] The evidence in [6], for example, indicates that the number of non-production workers employed is subject to less short-run variation than is the number of production workers employed.

Fifth, studies of investment decision should be aware of the different ways in which the cost of capital can affect investment expenditures and should attempt, whenever possible, to separate the effects of capital-labor substitution from other effects. Such a separation was attempted in [8] for the Cement and Steel industries. Also, the present model indicates that more attention should probably be given to the difference between replacement investment and investment in new capacity. Feldstein and Foot [10] have made an initial attempt in this direction. The above analysis indicates that replacement investment should be a function of the wage-rental ratio, and it would be of interest to see how important this effect is.

Most of the above suggestions require work with disaggregate data, and indeed one of the main points of the paper is that the use of seasonally adjusted quarterly data has tended to obscure much of the behavior of the firm. This does not necessarily imply, however, that the present model has no relevance for macro-economic models. The employment sector in the forecasting model in [9], for example, is based on the above model of employment decisions, and the results have been quite good. A measure of the aggregate amount of excess labor in the economy has been found to be significant in explaining the change in aggregate employment. Similarly, some of the ideas of the present model may be useful in the specification of aggregate inventory and investment equations. One might, for example, attempt to estimate the inventory equation in footnote 5 using aggregate data. In most macro models, the desired stock of inventories, V_t^d, is assumed to be a function only of the current level of sales, and what equation (6) and the related discussion suggest is that this specification is likely to be too simple. Consequently, some improvement in the specification of aggregate inventory equations might result from using some of the above ideas. Aggregate investment equations might also be improved by trying to distinguish among the various ways that the cost of capital can affect investment expenditures. This would probably require the use of employment and wage rate data as well as cost of capital data. Along these same lines, it might be fruitful to attempt to derive a measure of aggregate capacity and then separate the data on aggregate investment expenditures into data on capacity expenditures and data on replacement expenditures.

The type of work suggested here is for the most part work of a rather pedestrian nature and is to many people less exciting and less elegant than, say, deducing the behavior of the firm from a few basic postulates. It is the view here, however, that much work of this type is needed before one can hope to specify a realistic cost function and deduce the behavior of the firm from the minimization of this function. Perhaps it is appropriate to close with a quote from Francis Bacon made in 1605:

> For the wit and mind of man, if it work upon matter, which is the contemplation of the creatures of God, worketh according to the stuff and is limited thereby; but if it work upon itself, as the spider worketh his web, then it is endless, and brings forth indeed cobwebs of learning, admirable for the fineness of thread and work, but of no substance or profit.

REFERENCES

[1] Belsley, David A., Industry Production Behavior: The Order-Stock Distinction (Amsterdam: North-Holland Publishing Co., 1969).

[2] Bischoff, Charles W., "Lags in Fiscal and Monetary Impacts on Investment in Producers' Durable Equipment," Cowles Foundation Discussion Paper No. 250, June 11, 1968.

[3] Brechling, Frank, "The Relationship Between Output and Employment in British Manufacturing Industries," The Review of Economic Studies, XXXII (July 1965), 187-216.

[4] Childs, Gerald L., Unfilled Orders and Inventories: A Structural Analysis (Amsterdam: North-Holland Publishing Co., 1967).

[5] Coen, Robert M., and Bert G. Hickman, "Constrained Joint Estimation of Factor Demand and Production Functions," The Review of Economics and Statistics, LII (August 1970), 287-300.

[6] Fair, Ray C., The Short-Run Demand for Workers and Hours (Amsterdam: North-Holland Publishing Co., 1969).

[7] Fair, Ray C., "Sales Expectations and Short-Run Production Decisions," The Southern Economic Journal, XXXVII (January 1971), 267-275.

[8] Fair, Ray C., "Capital-Labor Substitution and the Demand for Capacity," Econometric Research Program, Research Memorandum No. 127, Princeton University, August 1971.

[9] Fair, Ray C., A Short-Run Forecasting Model of the United States Economy (Lexington: D. C. Heath and Company, 1971).

[10] Feldstein, Martin S., and David K. Foot, "The Other Half of Gross Investment: Replacement and Modernization Expenditures," The Review of Economics and Statistics, LIII (February 1971), 49-58.

[11] Griliches, Zvi, "Distributed Lags: A Survey", Econometrica, XXXV (January 1967), 16-49.

[12] Holt, Charles C., Franco Modigliani, John F. Muth, and Herbert A. Simon, Planning Production, Inventories, and Work Force (Englewood Cliffs, N.J.: Prentice-Hall, 1960).

[13] Jorgenson, Dale W., "Capital Theory and Investment Behavior," The American Economic Review, LIII (May 1963), 247-259.

[14] Kuh, Edwin, "Income Distribution and Employment Over the Business Cycle," in Duesenberry, et al., eds., The Brookings Quarterly Econometric Model of the United States (Chicago: Rand McNally, 1965).

[15] Lucas, R. E., Jr., "Optimal Investment Policy and the Flexible Accelerator," International Economic Review, VIII (February 1967), 78-85.

[16] Nadiri, M. Ishag, and Sherwin Rosen, "Interrelated Factor Demand Functions," The American Economic Review, LIX (September 1969), 457-471.

[17] Nerlove, Marc, "On Lags in Economic Behavior," Center for Mathematical Studies in Business and Economics, University of Chicago, Report 7109, February 1971.

[18] Salter, W.E.G., Productivity and Technical Change (Cambridge: Cambridge University Press, second edition, 1966).